RESEARCH REPORT

ISSN 1172-8531

Fast Kriging

R. K. Beatson and C. T. Mouat

Department of Mathematics & Statistics,
University of Canterbury,
Private Bag 4800, Christchurch, New Zealand.

Report Number: UCDMS2000/2 January 2(

Department of Mathematics and Statistics

Fast Kriging

R. K. Beatson and C. T. Mouat

Department of Mathematics and Statistics
University of Canterbury
Christchurch, New Zealand

April 3, 2000

Abstract

This paper presents a fast technique for fitting a Kriging surface of the form $s(\cdot) = \sum_j \alpha_j \Phi(\cdot - x_j) + \sum_{j=1}^{m} \gamma_j q_j(\cdot)$, where Φ is a semi-variogram determined from the data. Finding the coefficients of s by conventional techniques requires $\mathcal{O}(N^3)$ operations, and $\mathcal{O}(N^2)$ storage. Numerical evidence suggests that for typical Φ's the iterative method presented here requires $\mathcal{O}(N \log N)$ operations and $\mathcal{O}(N)$ storage.

Keywords: universal Kriging, preconditioned iterative method, moment method.

1 Introduction

Kriging is a surface fitting method which models spatial processes, $Z(x)$, in d dimensions. Areas where it has been applied include geophysics, hydrology, meteorology, mining engineering and bathymetry, see for example Cressie [9]. Kriging approximates the spatial process, $Z(x)$, under the assumption

$$Z(x) = \mu(x) + \kappa(x).$$

Here $\mu(x)$, the large scale variation, or the trend, of the process is an unknown element of some known finite dimensional subspace $\mathcal{P}$ of functions. $\kappa(x)$ is a zero mean intrinsically stationary random process. These assumptions on the variability of $Z(x)$ are described in detail in Cressie [10, §2.2.1 and §3.4]. When they hold the Kriging approximation minimizes the mean-squared prediction error

$$E\Big((Z(x) - s(x))^2\Big), \tag{1}$$

for each $x \in \mathcal{D} \subset \mathcal{R}^d$, where $Z(x)$ is the actual value of the surface and $s(x)$ is the Kriged value obtained using the linear unbiased predictor

$$s(x) = \sum_{i=1}^{N} \lambda_i z_i. \tag{2}$$

Here, $\{z_1, \ldots, z_N\}$ are the observed values at the corresponding spatial locations $\{x_1, \ldots, x_N\}$ and $\lambda = \lambda(x) = \{\lambda_1, \ldots, \lambda_N\} \in R^N$ is the vector of coefficients to be determined. These coefficients are subject to the constraints

$$\sum_{j=1}^{N} \lambda_j q(x_j) = q(x), \quad \text{for all } q \in \mathcal{P}, \tag{3}$$

which make the predictor unbiased. Let $m = \dim(\mathcal{P})$ and $\{q_1, \ldots, q_m\}$ be some basis for $\mathcal{P}$ then the constraints (3) can be written

$$\sum_{j=1}^{N} \lambda_j q_i(x_j) = q_i(x), \quad \text{for all } i = 1, \ldots, m. \tag{4}$$

Often $\mathcal{P}$ will be π_k^d, the space of polynomials of degree k in d variables. Then the constraints are the same as those found in the usual formulation of the radial basis function (RBF) interpolation equations. In the RBF setting they are interpreted as conditions which take away the extra degrees of freedom added by introducing the polynomial part. In the Kriging context they are conditions that make λ unbiased. When appropriate conditions (see equation (5)) on $Z(x)$ are met, minimizing the prediction error renders a surface which is as accurate as possible at the point x. The predictor in this case is called a best linear unbiased predictor. It results in a surface (2) that can be written in the RBF like form

$$s(\cdot) = \sum_{j} \alpha_j \Phi(\cdot - x_j) + \sum_{j=1}^{m} \gamma_j q_j(\cdot).$$

Furthermore, the system (14) and (16) that would be solved to find the coefficients in the above "dual" formulation of the Kriging surface, are identical with those of the usual formulation of the RBF interpolation problem.

As can be seen above, and is well known, Kriging and RBF fitting are highly related. Indeed for fixed Φ and polynomial degree the fits produced by both methods are identical. However, the motivations of the two methods, and the assumptions underlying them, are different. In RBF fitting, the choice of a quadratic smoothness penalty, also called an energy seminorm, will determine Φ and thus the RBF interpolant to be fitted. In the Kriging model, assumptions are made about the variability of the data as a function of position, and then Φ is determined experimentally from the data. Which method is more appropriate is application dependent. See [11, 12, 14, 15, 22] for discussions and numerical results comparing Kriging and splines.

Our discussion here applies to a type of Kriging known as universal Kriging with polynomial trend. This type of Kriging can be split into three parts: finding the semi-variogram (or basis function); forming the Kriged or fitted surface; and forming the prediction error surface. In this paper we concentrate on the second part, forming the Kriged surface.

The purpose of this paper is to present a fast method for forming the fitted surface when the number of data points, N, is large, say greater than 4000, and assuming the semi-variogram is already known. Experiments indicate that this method requires $\mathcal{O}(N \log N)$ operations and $\mathcal{O}(N)$ storage, whereas direct fitting requires $\mathcal{O}(N^3)$ operations and $\mathcal{O}(N^2)$ storage. The method used is a developement of a method which has been successful in the RBF context in the special cases of thin-plate spline and multiquadric Φ's. Here it is used instead with several semi-variograms commonly used in Kriging, and in combination with a moment based fast multiplication technique. Furthermore, the moment method, used as a fast evaluator, allows evaluation of the fitted surface at an incremental cost of $\mathcal{O}(1)$ operations per extra point.

The layout of the paper is as follows. Section 2 outlines the universal Kriging procedure. Section 3 describes a fast method based on preconditioned GMRES iteration, and moment method multiplication, for forming the Kriged surface. Section 4 presents numerical results using random data. Section 5 discusses forming the prediction error surface and Section 6 gives numerical results for a geophysical dataset.

We alert the reader that there are other iterative methods for fast solution of RBF interpolation systems that may also be useful in the Kriging context. We mention in particular an improved version of the method described in [4], which is currently under developement.

2 Surface fitting by universal Kriging

In this section we outline some of the theory of Kriging, and the "dual" form of the Kriging equations.

Universal Kriging with polynomial trend is a form of Kriging that is appropriate if the trend in the data can be approximated by an unknown polynomial of degree k, see Cressie [10, §3.4]. This type of Kriging can be used when $Z(x)$ satisfies,

$$\begin{aligned} \mu &\in \pi_k^d, \\ E(Z(x)) &= \mu(x), \quad \text{for all } x \in D, \\ \operatorname{var}(Z(x+h) - Z(x)) &= 2\Phi(h) \quad \text{for all } x, x+h \in D. \end{aligned} \tag{5}$$

The function $\Phi(h)$ is referred to as the semi-variogram and $2\Phi(h)$ as the variogram. If the semi-variogram is radial, i.e. $\Phi(h) = \phi(|h|)$, then $Z(x)$ is called isotropic, otherwise it is anisotropic. Often a simple linear transformation of the spatial locations can give a system which is isotropic in the new data.

The numerical results in this paper are only for the isotropic case although the method can easily be applied in the anisotropic setting. Due to microscale variation and measure-

ment error $\Phi(h)$ is often discontinuous at the origin. If this is the case we can write

$$\Phi(h) = \begin{cases} 0, & \|h\| = 0, \\ c_0 + f(h), & \|h\| \neq 0, \end{cases} \tag{6}$$

where $f(h)$ is not necessarily radial but is continuous at 0 with $f(0) = 0$. The equations specifying an universal Kriging surface are solvable whenever the semi-variogram, Φ, is strictly conditionally negative definite of order $k+1$ (SCND-$(k+1)$). See for example [4] for the definition of SCND-$(k+1)$ and a proof of solvability. In this paper we consider semi-variograms in $\mathcal{R}^2$ of type (6).

In the remainder of this section we develop the well known equations for finding the coefficients, $\lambda = \lambda(x)$, of the Kriged surface. A good source for these and other Kriging equations is [10]. Expanding the square term in equation (1), using the predictor (2), and constraints on λ (4), we obtain

$$\begin{aligned}
(Z(x) - s(x))^2 &= \left(Z(x) - \sum_i \lambda_i Z_i \right)^2, \\
&= \sum_i \lambda_i Z(x)^2 - 2 \sum_i \lambda_i Z_i Z(x) + \left(\sum_i \lambda_i Z_i \right)^2, \\
&= \sum_i \lambda_i (Z_i - Z(x))^2 - \sum_i \lambda_i Z_i^2 + \left(\sum_i \lambda_i Z_i \right)^2, \\
&= \sum_i \lambda_i (Z_i - Z(x))^2 - \sum_j \sum_i \lambda_j \lambda_i Z_i^2 + \sum_j \sum_i \lambda_j \lambda_i Z_j Z_i, \\
&= \sum_i \lambda_i (Z_i - Z(x))^2 - \frac{1}{2} \sum_j \sum_i \lambda_j \lambda_i (Z_i - Z_j)^2.
\end{aligned} \tag{7}$$

Now using equations (7) and (5) in (1) the following expression can be obtained for the prediction error

$$\begin{aligned}
E\Big((Z(x) - s(x))^2 \Big) &= E\left(\sum_i \lambda_i (Z_i - Z(x))^2 - \frac{1}{2} \sum_j \sum_i \lambda_j \lambda_i (Z_i - Z_j)^2 \right), \\
&= 2 \sum_i \lambda_i \Phi(x_i - x) - \sum_j \sum_i \lambda_i \lambda_j \Phi(x_i - x_j).
\end{aligned} \tag{8}$$

The coefficients $\lambda = \lambda(x) \in \mathcal{R}^N$ are found by solving a constrained minimization problem of the form

$$\begin{aligned}
&\min_\lambda \; E\Big((Z(x) - s(x))^2 \Big) \\
&\text{subject to } \sum_i \lambda_i q_j(x_i) = q_j(x), \quad j = 1, \ldots, m.
\end{aligned} \tag{9}$$

If we write $c_j(\lambda) = \sum_i \lambda_i q_j(x_i) - q_j(x)$ then the Lagrangian of this equality constrained problem can be written

$$l(\lambda, \nu) = E\Big((Z(x) - s(x))^2 \Big) - 2\sum_{j=1}^{m} \nu_j c_j(\lambda), \tag{10}$$

where $\nu = (\nu_1, \ldots, \nu_m)^T$ are Lagrange multipliers. The first order necessary conditions for a solution of (9) are then

$$\nabla_\lambda \, l(\lambda, \nu) = 0, \text{and} \quad \nabla_\nu l(\lambda, \nu) = 0. \tag{11}$$

Substituting (8) into the expression (10) we find that at a minimum

$$\begin{aligned} \sum_j \lambda_j \Phi(x_j - x_i) - \Phi(x_i - x) + \sum_{k=1}^{m} \nu_k q_k(x_i) = 0, \quad i = 1 \ldots N, \\ \sum_j \lambda_j q_i(x_j) = q_i(x), \quad i = 1, \ldots, m, \end{aligned} \tag{12}$$

which can be written in matrix form as

$$B_\Phi \begin{bmatrix} \lambda \\ \nu \end{bmatrix} = \begin{bmatrix} g \\ c \end{bmatrix}, \tag{13}$$

where

$$B_\Phi = \begin{bmatrix} A_\Phi & Q \\ Q^T & 0 \end{bmatrix}, \tag{14}$$

$$\begin{gathered} (A_\Phi)_{ij} = \Phi(x_i - x_j), \quad i, j = 1, \ldots, N, \\ Q_{ij} = q_j(x_i), \quad i = 1, \ldots, N, \quad j = 1, \ldots, m, \\ g = (\Phi(x - x_1), \ldots, \Phi(x - x_N))^T, \end{gathered}$$

and

$$c = (q_1(x), \ldots, q_m(x))^T.$$

Now, g is a function of the evaluation point x, so finding the surface will involve solving (13) for each evaluation point. This requires $\mathcal{O}(N^2)$ operations assuming N is small enough that we can form and store some suitable factorization of B_Φ, for example a $P^T LU$ or a QR factorization. Of course it will often be the case that N is too large for factorization of B_Φ to be practical. It is common practice [14, 15, 21] to find $s(x)$ using a subset consisting of the closest β points to x[2]. This approximate method involves something between $\mathcal{O}(\beta^2)$ and $\mathcal{O}(\beta^3)$ operations per point. Unfortunately, using this approach, minor

[2]For some data sets and some evaluation points the set of β points may not be unique. If this is the case then the set should be found in a consistent manner.

discontinuities can occur at evaluation points where the subsets of local indices change. Also, prediction errors and confidence intervals are larger. The following "dual" Kriging equations [21, 10, 16] suggest an alternative method to form the prediction surface. From (2) and (13) we obtain

$$s(x) = \begin{bmatrix} z^T & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \nu \end{bmatrix} = \begin{bmatrix} z^T & 0 \end{bmatrix} B_\Phi^{-1} \begin{bmatrix} g \\ c \end{bmatrix}. \tag{15}$$

Now B_Φ^{-1} is symmetric so we can solve

$$B_\Phi \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} = \begin{bmatrix} z \\ 0 \end{bmatrix}, \tag{16}$$

for $\begin{bmatrix} \alpha^T & \gamma^T \end{bmatrix}^T$ to get RBF-like coefficients that do not depend on x. Equation (15) then becomes

$$s(x) = \begin{bmatrix} \alpha^T & \gamma^T \end{bmatrix} \begin{bmatrix} g \\ c \end{bmatrix} = \sum_i \alpha_i \Phi(x - x_i) + \sum_{j=1}^{m} \gamma_j q_j(x). \tag{17}$$

Note that (17) and (16) are the usual expressions for an RBF, and the usual system used to determine the coefficients of an interpolatory RBF, respectively.

The possibility of using fast evaluation techniques similar to those developed for radial basis functions [3, 5] to reduce the operation count for finding the "dual" Kriging coefficients has been previously mentioned in passing by some authors including [14, 15, 7].

3 A fast fitting method for large N

In this section we present a method for forming the Kriged surface using the RBF-like coefficients in (17). The method involves a combination of a preconditioner, a fast matrix-vector multiplication code appropriate for the function Φ, and the GMRES iterative algorithm for solving linear systems. The numerical experiments of later sections will show that for several typical semi-variograms this method can determine the RBF-like coefficients of the Kriged surface in $\mathcal{O}(N \log N)$ operations.

A similar approach has been successfully applied to solving RBF interpolation problems in the special cases of thin-plate spline and multiquadric basic functions in [2]. In the current paper we use instead several semi-variograms Φ common in the Kriging context. Another difference here is that the matrix-vector products arising in the GMRES iteration are performed with a fast moment method [6, 1]. This is a method which computes the action of the matrix $A_\Phi = (\Phi(x_i - x_j))$ on a vector of coefficients in approximately $\mathcal{O}(N)$ operations and using only $\mathcal{O}(N)$ storage. Direct calculation of the same matrix-vector product requires $\mathcal{O}(N^2)$ operations and also $\mathcal{O}(N^2)$ storage, making large problems intractable. In contrast to the more established fast multipole like methods, the fast

moment method is highly adaptive to changes of basic function Φ. Changing to a different Φ requires only the coding of a one or two line function for the slow evaluation of Φ. This adaptivity makes the fast moment method well suited for the Kriging application where many different Φ's occur. The fast moment method can also be used to reduce the incremental cost of evaluation of the fitted surface, s, at a single additional point, x, to $\mathcal{O}(1)$ operations instead of $\mathcal{O}(N)$ operations.

The heuristic underlying the approximate Lagrange function preconditioning used is to change from the basis of functions, $\Phi_j = \Phi(\cdot - x_j)$, to a basis of functions ψ_j, where $\psi_j(x_i) \approx \delta_{ij}$. We write the new basis element, ψ_j, in the form

$$\psi_j(\cdot) = \sum_{k=1}^{m} c_{jk} q_k(\cdot) + \sum_{i=1}^{N} \theta_{ji} \Phi(\cdot - x_i). \tag{18}$$

Each ψ_j is constructed so that $\{\theta_{ji}\}_{i=1}^{N}$ is orthogonal to polynomials in the sense of equation (4).

This change of basis leads to the new system of fitting equations

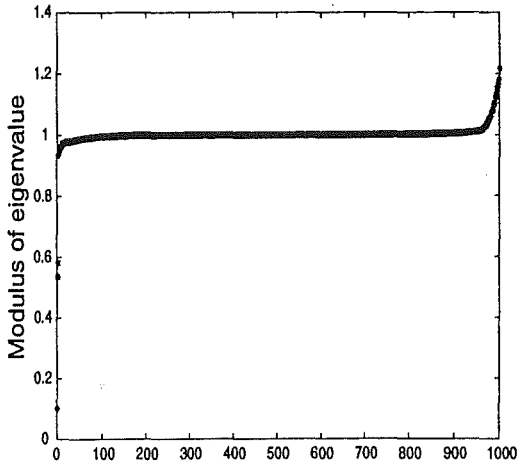
$$A_\psi y = z \tag{19}$$

where $(A_\psi)_{ij} = \psi_j(x_i)$. This is the linear system to which GMRES is applied. In practice A_ψ is never formed, as it is too expensive to store and use. Rather its action on a vector is calculated using the fast moment method in $\mathcal{O}(N)$ operations. To ensure fast convergence of the GMRES iteration we aim to choose the ψ_j elements so that A_ψ has eigenvalues that are clustered within a small relative radius. Such clustering is well known [13, 20, 8] to guarantee fast convergence of the GMRES iteration. Our ψ elements are constructed so that the ψ interpolation matrix has ones on the main diagonal and is close to zero everywhere else. Figure (1) shows the resulting clustered eigenvalues of A_ψ and the non-clustered eigenvalues of A_ϕ for the basis function $\phi(h) = h$.

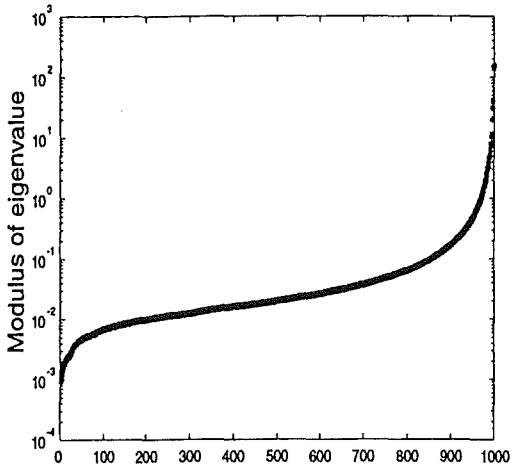
Different strategies for finding the θ_{ji}'s can greatly affect the performance of A_ψ in GMRES. If the θ_{ji}'s were chosen so that $\psi_j(x_j) = 1$ and $\psi_j(x_i) = 0$, $i \neq j$ then $A_\psi = I$ and GMRES would converge in one iteration. However, forming the ψ_j elements in this way would require the solution of N full size linear systems. Clearly this is not practical. To reduce the computation we restrict the number of non-zero $\{\theta_{ji}\}_{i=1}^{N}$ to $\beta << N$ for each j. We define the set $\mathcal{S}_j$ to be the set of β indices i such that θ_{ji} is possibly non-zero. Thus the new basis element is a sum over the indices in $\mathcal{S}_j$, or specifically

$$\psi_j(\cdot) = \sum_{k=1}^{m} c_{jk} q_k(\cdot) + \sum_{i \in \mathcal{S}_j} \theta_{ji} \Phi(\cdot - x_i). \tag{20}$$

Various strategies for finding the θ_{ji}'s are given in [2]. Some of these strategies are Φ specific and are unsuitable for this paper. The strategy we use for most of this paper is to pick the index set $\mathcal{S}_j$ as a set of indices of β closest points to x_j, together with the indices of a small number, τ, of special points. Then we form ψ_j by requiring that $\psi_j(x_j) = 1$ and



(a) Preconditioned matrix, A_ψ. Each $\mathcal{S}_j$ is fifty local points and nine special points.



(b) Unpreconditioned matrix, B_ϕ.

Figure 1: Eigenvalue plots for the basis function $\phi(h) = h$. The spatial data is one thousand random points within the domain $[0,1]^2$.

$\psi_j(x_i) = 0,\ i \in \mathcal{S}_j,\ i \neq j$. This strategy yields an approximate cardinal function, called a *local centres and special points approximate cardinal function* in [2]. The idea of including special points in $\mathcal{S}_j$ is to force ψ_j to zero at various points widely scattered throughout the domain. It is then expected that ψ_j will be close to zero near these points. If we were fitting within the square $[0,1]^2$ then a suitable choice of four special points would be the centres closest to $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$ respectively. Figure (2) shows a single new basis element formed with this strategy for the semi-variogram $\phi(h) = 1 - \exp(-h)$. It is clear from the graph that in this case the strategy has been extremely successful and the corresponding column of the matrix A_ψ will be very close to the j^{th} column of the identity.

Forming ψ_j involves solving a $(\beta+\tau+m)\times(\beta+\tau+m)$ system of interpolation equations. These systems can be converted to be symmetric positive definite using the method of [4]. Solving via Cholesky then takes $\mathcal{O}((\beta+\tau+m)^3/6)$ operations [3] To reduce the operation counts in this setup we often use the same Cholesky decomposition for more than one ψ element. In the examples given in table 1 about $0.1N$ Cholesky decompositions were formed and in table 2 about $0.3N$ Cholesky decompositions were formed. The preconditioning is slightly less effective using this technique due to the unbalanced nature of the subsets. In our experience the increase in GMRES iterations is only small and so forming ψ elements in this way is worthwhile.

Now to find the RBF-like coefficients $[\alpha^T\ \gamma^T]$ we simply convert from the good ψ basis

[3] We count operations in *old flops*, each old flop being one multiplication or division, together with one addition or subtraction, plus a little indexing.

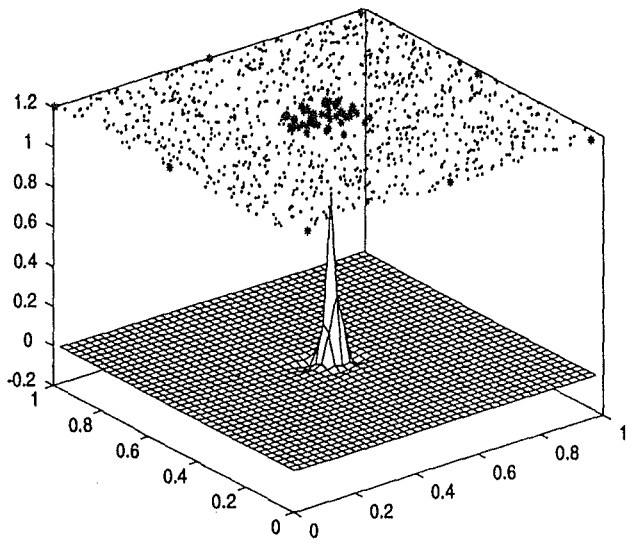


Figure 2: A new basis element based on fifty local points and nine special points out of a data set of one thousand points. The dots above the graph indicate the spatial data points and the asterix's indicate the location of points in $\mathcal{S}_j$.

to the bad Φ basis. Letting T be an $N \times N$ matrix with $T_{ij} = \theta_{ji}$ and C an $m \times N$ matrix with $C_{ij} = c_{ji}$ then equation (19) can be written

$$[A_\Phi \;\; Q] \begin{bmatrix} T \\ C \end{bmatrix} y = z. \tag{21}$$

The coefficients $[\alpha^T \;\; \gamma^T]$ can easily be found from

$$\begin{bmatrix} \alpha \\ \gamma \end{bmatrix} = \begin{bmatrix} T \\ C \end{bmatrix} y.$$

Exploiting the sparsity of T allows the above conversion from coefficients with respect to the good basis, to coefficients with respect to the bad basis, to be performed in $(\beta+\tau+m)N$ operations.

4 Numerical Results

This section presents numerical results generated with an initial implementation of the method of this paper. The method was applied to a selection of random data sets with various typical variograms valid in $\mathcal{R}^2$ and the computation times recorded. All the numerical examples are for the important special case of ordinary Kriging when the degree of the polynomial trend, k, is 0. The experiments were conducted on a Sun Ultrasparc machine. Section 6 below describes an application of the method to a non simulated data set, an electromagnetic survey.

All the semi-variograms considered are isotropic and of the form (6). The number of iterations for convergence of GMRES varies greatly depending on the preconditioning strategy used and also on the initial basis function. (22)-(25) specify $\phi(h)$ for $h > 0$, and we assume throughout that $\phi(0) = 0$.

exponential	$\phi(h) = c_0 + c_1(1 - \exp(-c_2 h))$	(22)
linear	$\phi(h) = c_0 + c_1 h$	(23)
power	$\phi(h) = c_0 + c_1 h^{c_2}, \quad c_2 \in [0, 2)$	(24)
rational quadratic	$\phi(h) = c_0 + c_1 \dfrac{h^2}{1 + h^2/c_2}$	(25)

All these functions are valid semi-variograms (SCND1 functions) provided $c_0, c_1, c_2 \geq 0$ and also $c_2 < 2$ in the case of (24) [10].

The results in tables 1 and 2 show that our approach can be successfully used for moderate and large problems. Many of the numerical experiments in these tables would not be solvable using standard direct methods. We have demonstrated that for twenty thousand centres the RBF-like coefficients can be found in less than two minutes in most cases. In the case of a linear variogram with each $\mathcal{S}_j$ consisting of one hundred closest points and nine special points the solution is found in 100.4 seconds. The choice of the size of $\mathcal{S}_j$ is a tradeof between minimizing the setup time and minimizing the time for convergence in GMRES. Increasing β would decrease the time for GMRES to converge but increase the setup time. For smaller values of N, say between one thousand and ten thousand, then having $\beta = 100$ rather than $\beta = 50$ has no clear advantage with respect to computation time. Once N is about twenty thousand a clear advantage can be seen in having larger subsets. However, larger subsets means an increase in storage requirements. The current computation times will be improved with algorithmic changes within both the moment method and the setup codes.

5 Prediction errors

One advantage of Kriging over RBF fitting is that Kriging is designed to minimize the prediction error at a point. To find the prediction error at a point x we solve the linear system in equation (13) for each evaluation point and then evaluate using equation (8). Clearly this is undesirable when the number of data points is large. In the RBF literature functions of the form of the right hand side of (8) are referred to as power functions. The fundamental properties of the power functions are now well known (see e.g. Schaback [23], Light and Wayne [17] and Powell [19]). The following result may be proved by an argument analogous to that of Light and Wayne [17, Lemma 2.7].

Theorem: Let $X_N = \{x_1, \ldots, x_N\}$ and $X_B = \{x_1, \ldots, x_B\}$ be finite subsets of $\mathcal{R}^d$ with $X_B \subset X_N$. For a given $x \in \mathcal{R}^d$, let $s_N(x)$ and $s_B(x)$ be the Kriged values formed using

Number of centres	Approximate cardinal function strategy	Iteration count to specified MSR error		Time in seconds for specified task	
		$< 10^{-6}$	$< 10^{-12}$	Setup	GMRES
4000	exponential	8	11	6.0	7.0
	linear	8	11	4.9	6.5
	power, c_2=1/2	9	11	5.8	6.9
	power, c_2=3/2	9	12	7.8	8.2
	rational quadratic	41	51	5.2	27.0
10000	exponential	13	17	17.2	26.0
	linear	13	16	14.0	24.4
	power, c_2=1/2	10	14	16.9	23.5
	power, c_2=3/2	10	15	22.5	26.1
	rational quadratic	56	67	14.6	93.1
20000	exponential	17	23	33.2	77.8
	linear	17	23	26.4	74.0
	power, c_2=1/2	16	21	32.1	71.7
	power, c_2=3/2	19	23	43.1	82.5
	rational quadratic	42	61	27.4	203.5

Table 1 Results of numerical experiments for Φ functions given by equations (22)-(25). The preconditioning elements consist of one hundred closest points and nine special points. GMRES timings are for 2-norm convergence to residual $< 10^{-6}$.

observations at the points of X_N and X_B respectively. Then

$$E\Big((Z(x) - s_N(x))^2\Big) \leq E\Big((Z(x) - s_B(x))^2\Big). \tag{26}$$

This result may be summarized as saying that using additional data points will not increase, and indeed is likely to decrease, the prediction error. Thus giving further motivation for fitting surfaces using all N points. If we fit using all N points and then form prediction errors based on local subsets of data we will obtain estimated prediction errors that are slightly greater than the actual prediction errors.

Figure 3 gives an example of prediction errors for a data set of 300 points uniformly distributed in $[0,1]^2$. This example shows the prediction error surface corresponding to using all points is only slightly below the prediction error surface corresponding to using local subsets of 30 points.

6 Geophysical application

This section describes the application of the method of this paper to a geophysical data set. The data considered is an electromagnetic survey consisting of measurements of radiation due to decay of uranium at 18824 spatial locations. The fittted surfaces are given in Figure

Number of centres	Approximate cardinal function strategy	Iteration count to specified MSR error		Time in seconds for specified task	
		$< 10^{-6}$	$< 10^{-12}$	Setup	GMRES
4000	exponential	13	17	3.6	9.3
	linear	13	17	2.7	9.4
	power, c_2=1/2	9	13	3.5	7.4
	power, c_2=3/2	10	15	5.2	9.5
	rational quadratic	31	46	2.8	24.3
10000	exponential	23	28	9.1	38.3
	linear	23	28	6.8	36.7
	power, c_2=1/2	15	22	8.7	31.3
	power, c_2=3/2	18	25	12.7	37.6
	rational quadratic	46	64	7.2	85.5
20000	exponential	29	35	18.2	107.2
	linear	29	34	13.5	101.3
	power, c_2=1/2	25	32	17.6	99.3
	power, c_2=3/2	29	35	25.5	112.6
	rational quadratic	34	62	14.3	203.1

Table 2: Results of numerical experiments for Φ functions given by equations (22)-(25). The preconditioning elements consist of fifty closest points and nine special points. GMRES timings are for 2-norm convergence to residual $< 10^{-6}$.

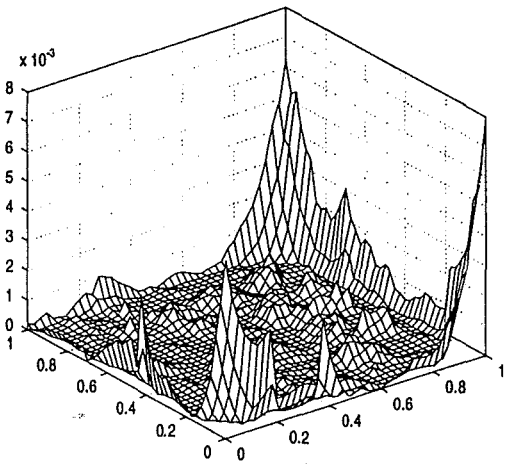
5. We are grateful to the Australian Geological Survey Organisation for the use of this data. We have scaled the data so it is contained in the region $[0, 0.5] \times [0, 1]$. The measurements of uranium radioactivity were taken by an aeroplane flying in transects across the domain. Assuming stationarity for this example we fitted a variogram by standard parametric techniques (see [24, 18] for more detail on fitting variograms). The experimental semi-variogram and fitted semi-variogram can be seen in Figure 4.

The fitted equation is

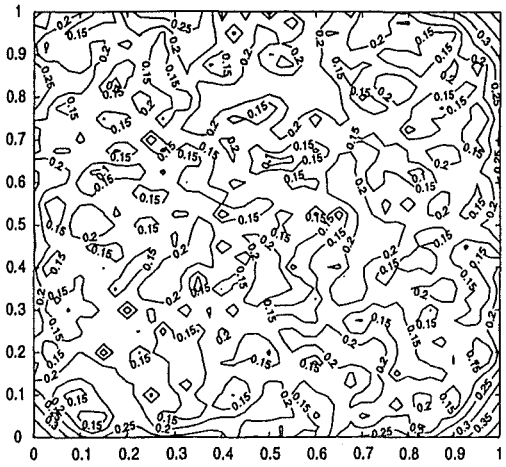
$$\phi(h) = 0.014 + 0.025(1 - \exp(-19h)), \quad \|h\| > 0, \tag{27}$$

which is a valid semi-variogram, that is an SCND1 function, in $\mathcal{R}^2$.

Our preconditioning approach here was to use only nearby centers and no special points in forming the ψ functions. ψ elements formed in this way are called *pure local approximate cardinal functions* in [2]. Using this preconditioner we were able to make our subset size 30 for each new basis element. This decreased our setup time considerably. Convergence to MSR 10^{-6} and MSR 10^{-12} took 11 and 18 iterations respectively. This compares favourably with the 17 and 23 iterations seen for the random data set of twenty thousand points in table 1. The total set up time was 11.8 seconds and GMRES iteration required a further 84.5 seconds for convergence of the 2-norm residual to 10^{-6}. Solving for the RBF coefficients by standard methods would take hours of computer time.



(a) Relative difference between the standard errors based on the whole data set and standard errors based on subsets of thirty points.



(b) Standard error surface where each point is found using the entire data set.

Figure 3: Standard error surfaces for the basis function $\phi(h) = 1 - \exp(-h)$. The spatial data is three hundred random points within the domain $[0, 1]^2$.

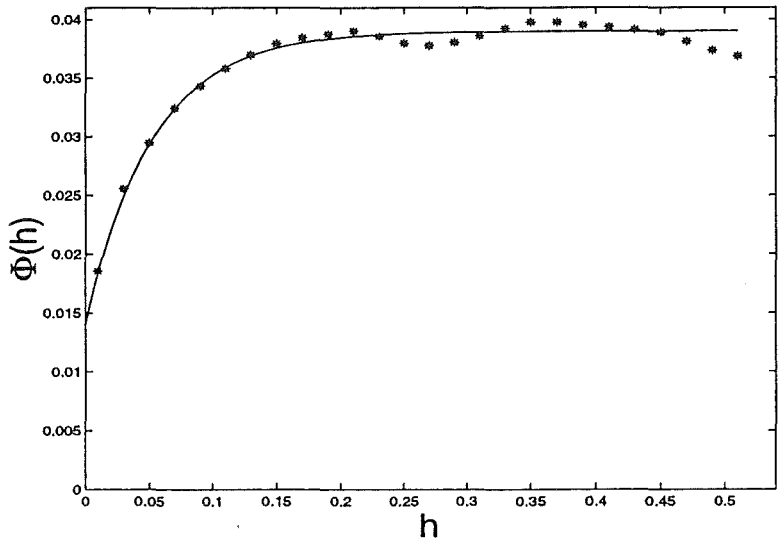
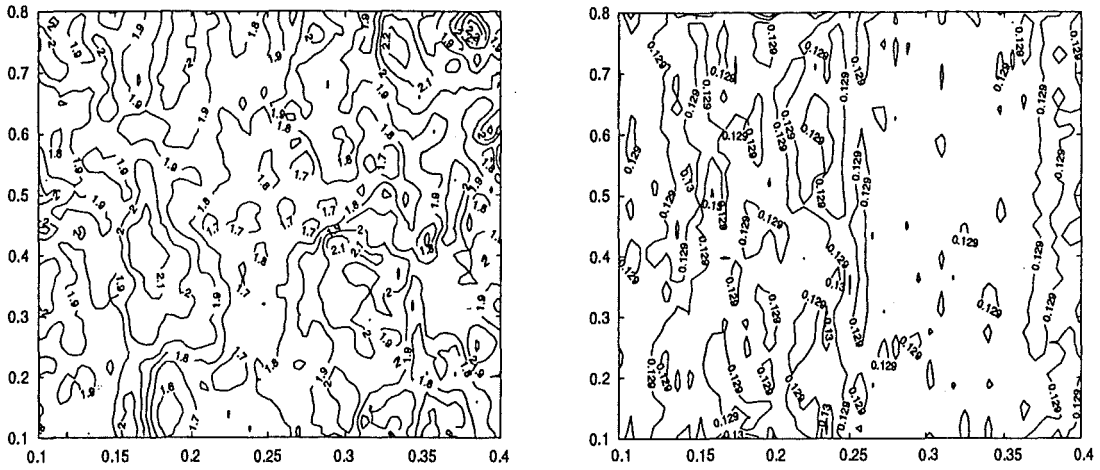


Figure 4: Experimental (*) and fitted (-) semi-variograms for the geophysical data set. The fitted semi-variogram is given by equation (27).

7 Discussion

We have presented a fast method for forming the fitted Kriging surface using RBF-like coefficients. In numerical experiments with the method fitting takes $\mathcal{O}(N \log N)$ operations



(a) Kriged surface of radiation due to decay of uranium over the scaled area.

(b) Standard error surface where each point is found using subsets of size 40

Figure 5: The Kriged surface and the prediction error surface for the geophysical uranium data. The semi-variogram is given by equation (27).

and $\mathcal{O}(N)$ storage. Previously finding these coefficients would have required $\mathcal{O}(N^3)$ operations and $\mathcal{O}(N^2)$ storage, therefore making the use of "global" Kriging surfaces impossible for large data sets. The numerical experiments presented show the effectiveness of this new method for a number of isotropic variogram functions using simulated data, and also for a geophysical data set. It is now possible to find these fitting coefficients for a data set containing 20,000 points in less than 2 minutes. Planned improvements in several aspects of the numerical code should extend the size of data set that can be handled to hundreds of thousands, or millions, of points in the near future.

References

[1] R. K. Beatson and E. Chacko, Fast evaluation of radial basis functions: A multivariate momentary evaluation scheme, to appear in *Curve and Surface Fitting: Saint Malo 1999*, A. Cohen, C. Rabut and L.L. Schumaker (eds), Vanderbilt Univ. Press, Nashville, 2000.

[2] R. K. Beatson, J. B. Cherrie and C. T. Mouat, Fast fitting of radial basis functions: Methods based on preconditioned GMRES iteration, *Advances in Computational Mathematics*, **11** (1999), 253-270.

[3] R. K. Beatson and W. A. Light, Fast evaluation of radial basis functions: Methods for 2-dimensional polyharmonic splines, *IMA Journal of Numerical Analysis*, **17** (1997), 343-372.

[4] R. K. Beatson, W. A. Light and S. Billings, Fast solution of the radial basis function interpolation equations: Domain decomposition methods. Manuscript, 1999.

[5] R. K. Beatson and G. N. Newsam, Fast evaluation of radial basis functions: I, *Computers and Mathematics with Applications*, **24**, **No. 12** (1992), 7-19.

[6] R. K. Beatson and G. N. Newsam, Fast evaluation of radial basis functions: moment-based methods, *SIAM J. Sci. Comput.*, **19** (1998), 1428-1449.

[7] S. D. Billings, G. N. Newsam and R. K. Beatson, Arbitrary basis functions: A general framework for interpolating and smoothly fitting geophysical data. Manuscript.

[8] S. L. Campbell, I. C. F. Ipsen, C. T. Kelley and C. D. Meyer, GMRES and the minimal polynomial, *BIT*, **36** (1996), 664-675.

[9] N. Cressie, The origins of Kriging, *Mathematical Geology*, **22** (1990), 239-252.

[10] N. Cressie, *Statistics for Spatial Data*, John Wiley and sons, inc., 1991.

[11] O. Dubrule, Two methods with different objectives: Splines and Kriging, *Mathematical Geology*, **15**(1983), 245-257.

[12] O. Dubrule, Comparing splines and Kriging, *Computers and Geosciences*, **10** (1984), 327-338.

[13] C. T. Kelley, *Iterative Methods for Linear and Non-linear Equations*, Society for Industrial and Applied Mathematics (1995).

[14] M. F. Hutchinson, On thin plate splines and Kriging, *Computing Science and Statistics*, **25** (1993), 55-62.

[15] M. F. Hutchinson and P. E. Gessler, Splines - more than just a smooth interpolator, *Geoderma*, **62** (1994), 45-67.

[16] D. Y. Le Roux, C. A. Lin and A. Staniforth, An accurate interpolating scheme for semi-Lagragian advection on an unstructured mesh for ocean modelling, *Tellus* **49A**(1997), 119-138.

[17] W. Light and H. Wayne, On power functions and error estimates for radial basis function interpolation, *Journal of Approximation Theory*, **92** (1998), 245-266.

[18] A. B. McBratney and R. Webster, Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates, *Journal of Soil Science* **37** (1986), 617-639.

[19] M. J. D. Powell, The uniform convergence of thin plate spline interpolation in two dimensions, *Numerische Mathematik*, **68** (1994), 107-128.

[20] Y. Saad and M. H. Schultz, GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear equations, *SIAM Journal Sci. Stat. Comp.*, **7** (1986), 856-869.

[21] F. Trochu, A contouring program based on dual Kriging interpolation, *Engineering and Computers*, **9** (1993), 160-177.

[22] G. S. Watson, Smoothing and interpolation by Kriging and with splines, *Mathematical Geology*, **16** (1984), 601-615.

[23] Z. Wu and R. Schaback, Local error estimates for radial basis function interpolation of scattered data, *IMA Journal of Numerical Analysis*, **13** (1993), 13-27.

[24] D. L. Zimmerman and M. B. Zimmerman, A comparison of spatial semivariogram estimators and corresponding ordinary Kriging predictors, *Technometrics*, **33** (1991), 77-91.